

基於混合注意力機制與長短期記憶之股票趨勢預測

Combining Hybrid Attention Networks and LSTM for Stock Trend Prediction

劉馨文 Hsin-Wen Liu

國立臺北科技大學資訊工程系

Department of Computer Science and Information Engineering

National Taipei University of Technology

t107598027@ntut.org.tw

王正豪 Jenq-Haur Wang

國立臺北科技大學資訊工程學系

Department of Computer Science and Information Engineering

National Taipei University of Technology

jhwang@ntut.edu.tw

摘要

本研究結合長短期記憶(LSTM)中股價時間序列的特徵，以及混合注意力模型(HAN)中模擬人判讀新聞影響股票重要性，有效地學習股價時間序列和新聞報導中的訊息順序，藉此訓練新聞文章與歷史股票交易資料之間的關聯，建構股票漲跌趨勢之模型。根據實驗結果，相較於僅單獨使用新聞或是股價資訊的模型，加入時間序列的新聞文章訊息使模型更能精準的預測股價市場趨勢，在兩種資訊的結合中，與 HAN、LSTM 模型相比，最佳準確度為 80%，整體最高可提升 40%的準確度。

Abstract

Our research merge two different models of the Hybrid Attention Networks (HAN) and the Long Short-Term Memory (LSTM) to improve the stock trend prediction. The combination of the two algorithms helps leverage the advantages of both models to learn sequential information in time series and news articles. The experimental results show that the best accuracy score, combined with news and stock prices, is 80 %. The performance of the proposed model compared to HAN and LSTM model increased by up to 40%.

關鍵詞：深度神經網路、注意力機制、股票預測

Keywords: deep attention networks, attention mechanism, stock prediction

一、緒論

股票市場一直是非常熱門且具有挑戰性，而一般投資人往往參考財經相關新聞或是技術指標來評估投資標的，而隨著科技的進步，網路時代的發達，電子媒體則是網路族群獲得資訊的主要來源，當重大的新聞消息發佈時，往往衝擊投資人的交易行為，造成股價一定程度的波動。因此，面對各式各樣的消息，投資者在投資時的規劃與預測分析是相當重要。

目前對股票趨勢預測的方法主要有：

- 利用時間序列中一段股票歷史數據
- 利用自然語言處理技術，金融新聞能夠有效地影響股票價格

但是股票歷史數據與新聞文本之間的結合與研究相對較少，大部分考慮單一新聞或是股價資訊影響，因此本研究將探討如何融合兩種不同的資料來源，並且使用深度學習的方式，藉此訓練兩者之間的關係，提高股票預測的準確度。

二、相關研究

（一）．基於股價之股票預測

股票市場每天都會產生大量的交易數據，提供大量資料，有利於深度神經網路訓練以及提高其預測能力。Hiransha M 等人[1]提出使用不同的深度學習網路，像是多層感知器(MLP)模型、卷積神經網路(CNN)、遞歸神經網路(RNN)和長短期記憶神經網路(LSTM)來比較各個模型之間的表現差異，並且觀察到神經網路模型優於時間序列的線性模型。Kai 等人[2]根據股票的歷史資訊結合 LSTM 神經網路，來預測後三天的股票趨勢是屬於哪一個區間，與隨機預測方法相比，LSTM 模型將股票收益預測的準確性從 14.3% 提高到 27.2%。

許多研究發現，深度學習神經網路具有儲存能力，因此它比其他機器學習方法

有更好時間序列數據的學習能力。

（二）．基於新聞之股票預測

隨著自然語言處理(Natural Language Processing) 的發展，使電腦把輸入的語言變成有意思的符號和關係，根據其目的再處理。Heeyoung Lee 等人[3]實驗有包含文字或財經訊息的模型，發現文字，是影響股票預測的重要性。自然語言處理(NLP)中單詞嵌入的發展，深度神經網路可以通過學習單詞向量來有效地掌握文本中的訊息。Xiao Ding 等人[4]使用 Open Information Extraction techniques(Open IE)提取事件 Event Extraction 並且用 WordNet 和 VerbNet 概括結構化事件特徵，以減少其稀疏性，Xiao Ding 等人[5]提出該方法的延伸，使用神經張量網路(Neural Tensor Network)來訓練新聞標題的事件嵌入，並使用卷積神經網絡(CNN)來預測 S&P 500 及其成分股的波動率。Hu Liu 等人[6]用新聞向量和價格數據訓練雙向 GRU 模型來預測股票的日波動率。他們的結果發現，市場新聞和股票價格的綜合訊息可以提高在日內交易上下文中對股票期貨價格回報預測的準確性。

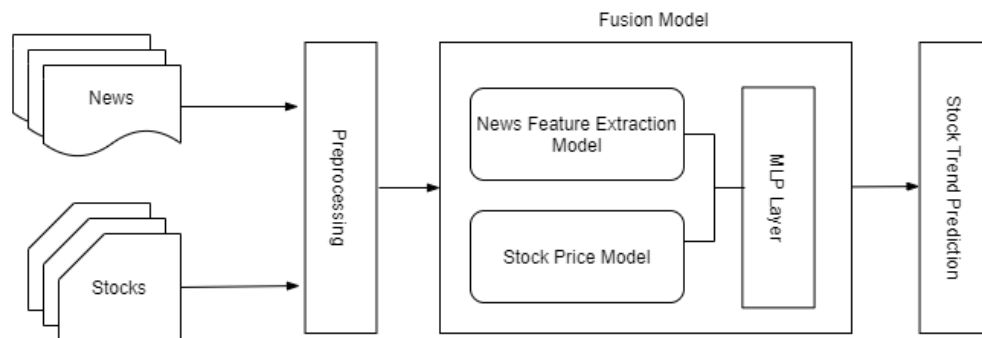
（三）．基於結合新聞和股價之股票預測

Yuzheng Zhai 等人[7]使用支援向量機(Support Vector Machine)的方式，提出結合相關的新聞和技術指標可增強股票趨勢的可預測性，實現更高的精準度。Xiaodong Li 等人[8]使用深度學習的方式，結合新聞文章的情緒、技術指標和股價，輸入於 LSTM 進行訓練和預測，此篇論文證實，包含新聞與股價的模型優於僅使用單一新聞或是股價資訊的模型，並發現於四個情緒字典中，特定於金融領域的情緒字典(Loughran – McDonald Financial Dictionary)對新聞情緒進行了更好的建模，與其他三個詞典相比，可以更好地提高預測性能。

總結先前的方法，我們知道用來預測股票的趨勢可以分為三大類：股價、新聞、以及結合新聞與股價。本研究中，我們將嘗試結合新聞文本與股價以進行股票之預測，透過結合兩種資料的序列特徵，加強模型對股票趨勢預測之精準度。

三、研究方法

我們在此章節說明研究的方法及架構，本研究主要分為四大部分，分別是系統架構、資料前處理、融合模型之選擇、模型融合之方法。



圖一、系統架構圖

（一）系統架構

Preprocessing 負責處理新聞文章資料集、歷年股價資料集、融合模型資料之產生。**Fusion Model** 負責結合兩種模型，分別為針對新聞文章進行訓練與針對股價進行訓練，並且把 **Fusion Model** 訓練出來之特徵矩陣輸入於多層感知器

（**MLP**）訓練。**Stock Trend Prediction** 負責把多層感知器(**MLP**)輸出之向量轉為漲、跌或是持平。

（二）資料前處理

1. 新聞資料前處理

標準普爾 500(**S&P 500**)為在美國股市的兩大股票交易市場。我們在處理新聞資料時，會以此 500 間公司當成目標，從新聞資料集中找出提到這幾間公司的文章。

首先我們會過濾掉標題和內容沒有出現於 **S&P 500** 公司的新聞，並且根據新聞發布的日期時間序列排序。根據 Jey Han Lau 等人[9]證實長文本中 **Doc2Vec** 的效果優於 **Word2Vec**，因此本論文中採用 **Doc2Vec** 的 **PV-DM** 模型來訓練文本向量。**PV-DM** 模型保留文檔中的單詞順序，經有轉換後新聞的格式為每一筆

[5,40,200]的三維矩陣，在三維矩陣中 5 代表天數，40 代表一天最多有 40 篇文章，200 代表文章向量維度。

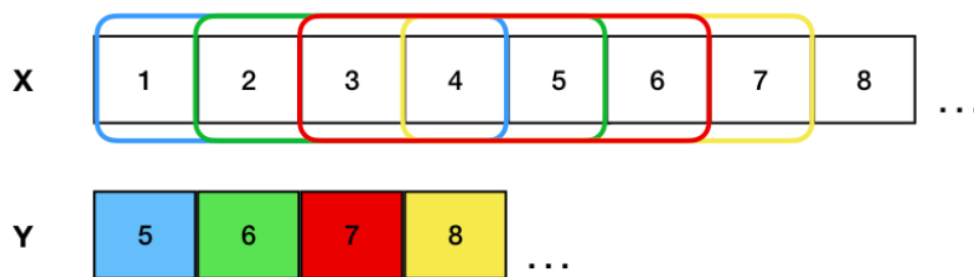
2. 股價資料前處理

我們將股價按照時間序列排序，並經由 Z 分數(Z-Score)標準化值，公式如式

(1)，資料將符合標準常態分佈(Standard Normal Distribution)，可透過 Z 分數標準化來降低離群值對整個模型的影響。資料經過標準化後，它能帶給模型兩個優點，提升模型的收斂程度、提高模型的精準度。

$$Z = \frac{X - \mu}{\sigma}, \sigma \neq 0 \quad (1)$$

為了輸入於長短期記憶神經網路(LSTM)模型，我們必須經過以下轉換格式，假設給定時間序列 N 天的長度，預測第 N+i 天的股票趨勢 Yi 為漲、跌或是持平，時間從 i 到 i + N - 1 天的股價資料表示為 X_i : $[X_i, \dots, X_{i+N-1}]$ ，預測的資料為 Y_i 而 (X_i, Y_i) 被用作 LSTM 神經網路的輸入，如圖二[12]所示，window size 長度為 4 天的滑動窗口示意圖。



Sliding window algorithm of sequence length 4, for data (X) and corresponding labels (Y).

圖 二 、模型輸入示意圖

3. Fusion Model 資料前處理

在上面小節我們有介紹到，針對新聞文章為了降低 Model 訓練時的記憶體使用量，我們使用 Doc2Vec 將複雜文章轉換為 200 維度之矩陣，針對股價為了避免各家公司股價高低不同造成模型預測失準，我們使用 Z 分數標準化(Z-Score

Standardization)來正規化股價資訊。

Fusion Model 由於是結合兩種不一樣之模型，因此輸入也必須做處理才能使 Fusion Model 進行訓練，我們必須串接文章向量與股價資訊，而我們為了要讓模型能夠同時拿到新聞文章 200 維度之特徵和股價向量開盤、收盤之特徵，我們使用串接(Concatenate)的方式結合兩種資料，讓產生出來的資料集每一筆都包含新聞文章與股價之特徵。

4. 股價標籤分類

股價趨勢預測通常有上漲、持平、下跌三個級距，以下介紹歸類之方法。對於給定日期 t 和股價 s ，參考 Ziniu Hu 等人[6]對於股票標籤分類的評估方法，通過以下公式計算其上漲百分比，公式如式(2)。把當日開盤價格與前一天的開盤價格相減，除以前一天開盤價格，可以計算出其上漲的百分比幅度，並依照其上漲百分比歸為不同的類別。

$$Rise_Percent_{(t)} = \frac{Open_Price_{(t+1)} - Open_Price_{(t)}}{Open_Price_{(t)}} \quad (2)$$

為了使類別的判斷能夠平均分布，因此定義閾值來幫助類別判斷之篩選，經過實驗後，發現 0.45% 閾值，能使三個類別(漲、跌、持平)的資料集數量平均分布。

- 漲(Up):
 - $Rise_Percent(t) > 0.45\%$
- 跌(Down):
 - $Rise_Percent(t) < -0.45\%$
- 持平(Preserve):
 - $-0.45\% \leq Rise_Percent(t) \leq 0.45\%$

公式(3)為 S&P 500 所有公司在 2016~2017 年期間，股價趨勢為漲的資料集總數、公式(4)為 S&P 500 所有公司在 2016~2017 年期間，股價趨勢為跌的資料

集總數、公式(5)為 S&P 500 所有公司在 2016~2017 年期間，股價趨勢為持平的資料集總數， n 為 S&P 500 公司總數量、 $threshold$ 為不同 $Rise_Percent$ 參數。

$$total_{Up} = \sum_{i=1}^n Up_i^{threshold} \quad (3)$$

$$total_{Down} = \sum_{i=1}^n Down_i^{threshold} \quad (4)$$

$$total_{Preserve} = \sum_{i=1}^n Preserve_i^{threshold} \quad (5)$$

公式(6)為 S&P 500 所有公司在 2016~2017 年期間，股價趨勢資料集總數量、公式(7)是股價趨勢為漲的資料集在全部資料集中所佔的百分比、公式(8)是股價趨勢為跌的資料集在全部資料集中所佔的百分比、公式(9)是股價趨勢為持平的資料集在全部資料集中所佔的百分比。如表 3.1 顯示，當 $threshold (Rise_Percent)$ 為 0.45%時，能使股價趨勢為漲、跌、持平的資料集在全部資料集中所佔的百分比趨近於相等。

$$total = total_{Up} + total_{Down} + total_{Preserve} \quad (6)$$

$$Distrubition_Percent_{Up} = \frac{total_{Up}}{total} \quad (7)$$

$$Distrubition_Percent_{Down} = \frac{total_{Down}}{total} \quad (8)$$

$$Distrubition_Percent_{Preserve} = \frac{total_{Preserve}}{total} \quad (9)$$

(三) 模型融合之選擇

本章節將介紹新聞與股價模型之選擇，以及其選擇的原因。

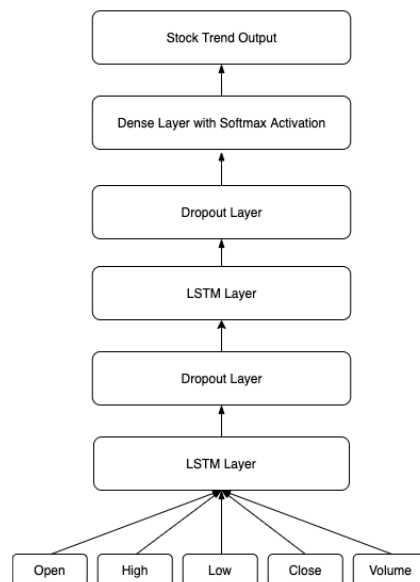
1. 混合注意力機制(Hybrid Attention Networks)

由於 Zichao Yang 等人[10]提出了 Hierarchical Attention 方式去進行文本分類，相較於先前研究於文本分類模型中是最好的，Ziniu Hu 等人[6]提出的混合注意力機制(Hybrid Attention Networks)模型，是根據 Zichao Yang 等人[10]提出的 Hierarchical Attention 架構應用於股票趨勢預測，賦予文章不同的權重，找出影

響股票的重要資訊，經由實驗證實效果為最佳。

在新聞文章訓練時，我們會需要各個文章彼此之間的關聯性而不是每篇文章獨立去進行訓練，因此我們需要選擇一種模型來幫助我們達到訓練文章彼此之間的關聯性，針對影響股價的重要性，賦予新聞文章不同的權重，使模型有效地學習人在閱讀新聞文章中的順序訊息，能夠幫助我們更精準的達到股票趨勢預測之效果。

2. 階層式長短期記憶神經網路(Hierarchical LSTM)



圖三、Hierarchical LSTM 架構圖

本篇論文架構採用長短期記憶網路(LSTM)，參考 Kai 等人[2]的設計，為了在順序訊息中找出時間序列的特徵，加上 LSTM 具有存儲能力，可以幫助改善預測效能，因此本研究選擇 LSTM 模型來訓練股價。本研究所提出的階層式長短期記憶神經網路(Hierarchical LSTM)架構中，經過實驗，發現兩層的 LSTM 效果最佳。為了避免過度擬合，每層 LSTM 連接一個 Dropout Layer 於每一個序列時間長度中，Dropout 將會根據訓練中的給定機率分佈將某些向量設置為零，對於遞迴神經網路，Dropout 可以迫使循環層更穩健地執行其中間運算，而不會移除單元狀態中的所有訊息。

(四) 模型融合之方法

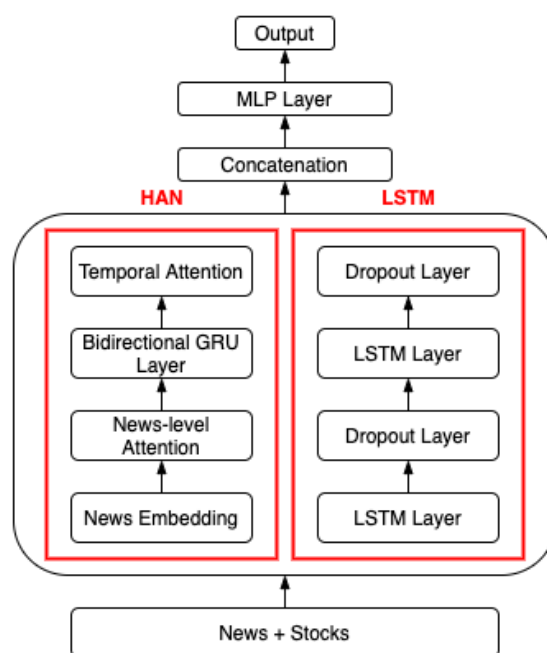


圖 四 、 Fusion Model 架構圖

Fusion Model 結合方式如圖四。使用兩個模型完整的向量輸出，結合 (concatenate) 時間序列注意機制 (Temporal Attention) 與 LSTM 模型架構中第二層 LSTM 輸出最後一個隱藏狀態作為輸出能達到以下效果，Temporal Attention 能夠在時間序列的新聞文章當中找出影響力最大的新聞文章子集合、透過使用兩層 LSTM 和 Dropout layer 的結合能夠使股票預測之效果更為精確方便找出股價的重要特徵，結合後再輸入於 MLP 進行分類來達到使用新聞文章與股價進行股票趨勢預測之效果。時間序列注意機制 (Temporal Attention) 設計的位置也會對模型造成一定的影響，在 Fusion Model 中只針對新聞文章進行注意力權重訓練，從新聞文章中找出影響力較大的文章跟 LSTM 輸出的股票預測向量結合，以此達到透過影響力大之新聞文章來影響股價預測之效果。

透過混合注意力機制 (HAN) 模型強化判讀新聞影響股票重要性、長短期記憶神經網路 (LSTM) 學習時間序列的特徵，使模型更有效的結合兩種不同類型的資

訊，提高預測的精準度。

四、實驗與討論

(一)．資料集

本研究所使用的資料集為 Kaggle 上所提供的路透(Reuters)財經新聞文章、標準普爾 500(S&P 500)的股價歷史資訊。如表一所示，分別為新聞文章數量以及股價公司資料筆數。資料的處理方式為 80%訓練集，20%測試集。

表 一、新聞文章與股價的資料數量

	Reuters	S&P 500
數量	10686 篇	500 間
時間	2016/1/1~2017/7/5	2016/1/1~2017/7/5

(二)．環境設置

研究使用 Adam[11]作為優化器，並設定其初始學習率為 5×10^{-5} ，每一次訓練都使用 Early Stopping，而 Early Stopping 的設定為 `monitor = 'val_loss'`, `min_delta = 0.0001`, `mode = 'min'`, `verbose = 2`, `patience = 200`，避免過度擬合、學習率過大導致不收斂。Batch size 為 8，最大 epoch 數為 500。使用 Three-fold-cross-validation 來評估模型之精準度。

(三)．HAN 模型實驗

我們針對混合注意力機制(HAN)模型進行實驗，根據之前的相關研究以 Window Size 為 5 天預測第 6 天和 window size 為 10 天來預測第 11 天的股票趨勢，來探討是否影響股票趨勢預測。預測新聞文章中出現過的標準普爾 500(S&P 500)公司之股票趨勢準確率，並根據全部準確率加總平均，由表二可以發現，長(Window Size 為 10)和短(Window Size 為 5)，準確度皆為 0.4，對於混合注意力機制(HAN)模型沒有太大的差異和影響。

表 二、HAN 模型在不同 Window size 之比較圖

	Window Size = 5	Window Size = 10
Average Accuracy	0.4	0.4
Average F1 Score	0.33	0.31

(四) . Hierarchical LSTM 模型實驗

本小節將探討，Window Size 長短對於階層式長短期記憶神經網路(Hierarchical LSTM)模型之影響以及股票特徵選取之比較。針對標準普爾 500(S&P 500)每一間公司算出各自的準確率並且加總平均。由表三可以發現對於時間長短之影響，其中又以短期(Window Size 為 5)的時間序列效果較好。針對特徵之選取，相較於其他特徵的選取，單獨只取收盤(Close)特徵的股票趨勢預測較佳。根據實驗結果，可以發現 LSTM 適合處理和預測時間序列的問題，對於短期的預測效果較佳，選取單一特徵，將不受其他資訊的干擾，可以有更好的表現。

表 三、Hierarchical LSTM Model 不同 Window Size 比較圖

Window Size = 5	Close	Close/Volume	Close/Open/High/Low /Volume
Average Accuracy	0.72	0.46	0.65
Average F1 Score	0.7	0.41	0.63
Window Size = 10	Close	Close/Volume	Close/Open/High/Low /Volume
Average Accuracy	0.68	0.42	0.55
Average F1 Score	0.56	0.39	0.51

(五) . Fusion Model 實驗

為了探討 Fusion Model 的有效性，我們討論時間序列對於模型的影響以及特徵之選取的比較，針對標準普爾 500(S&P 500)每一間公司算出各自的準確率並且加總平均。由表四可以發現，本篇論文所提出的方法架構，結合新聞和股價，

對於整體股票預測趨勢有顯著的提升，其中，又以短期(Window Size 為 5)預測效果較長期(Window Size 為 10)為佳，特徵選去的部分，則是拿單一收盤(Close)的準確率最佳。

表 四、Fusion Model 不同 Window Size 比較圖

Window Size = 5	Close	Close/Volume	Close/Open/High/Low /Volume
Average Accuracy	0.8	0.61	0.71
Average F1 Score	0.79	0.57	0.69
Window Size = 10	Close	Close/Volume	Close/Open/High/Low /Volume
Average Accuracy	0.77	0.47	0.62
Average F1 Score	0.71	0.45	0.57

(六)．模型加入股價資訊的效果

本小節針對 Ziniu Hu 等人[6]提出混合注意力機制(HAN)模型相比，此論文的 Window Size 設定為 10，因此我們也把 Window Size 設定為 10，輸入我們的資料集以及比較本研究所提出的混合深度模型(Fusion Model)，綜合比較不同模型之間的效果。如表五。可以觀察到本篇論文所提出的架構，加入股價時間資訊，對於整體有非常大的提升。原本只有新聞的混合注意力機制(HAN)模型從 0.4，加入股價後，準確率可以達到 0.77。

表 五、加入股價資訊的效果比較

Model	HAN With Reuters	Fusion Model With Reuters + S&P 500
Average Accuracy	0.4	0.77
Average F1 Score	0.33	0.71

(七)．新聞與股價資訊結合方法的效果

表 六、各個模型綜合比較表

	HAN	Hierarchical LSTM	Fusion Model
Average Accuracy	0.4	0.72	0.8
Average F1 Score	0.33	0.7	0.79

由表六中，我們可以看到本論文提出的 Fusion Model 準確度(80%)在所有模型當中表現最佳，Fusion Model 於最後結合新聞與股價有助於提升模型整體的效果，因為不同模型能夠針對不同資料分別擷取更有效的特徵。

五、結論

本研究提出的 Fusion Model，此模型結合新聞與股價資料集，新聞採用混合注意力機制(HAN)模型，模擬人判讀新聞影響股票重要性，股價方面則是採用長短期記憶神經網路(LSTM)模型，擷取股價時間序列的特徵，有效地學習新聞報導中的訊息順序和股價時間序列，藉此訓練新聞文章與歷史股票交易資料之間的關聯，建構股票漲跌趨勢之模型。由實驗結果可以證實，結合新聞與股價的混合深度模型優於單僅使用新聞或是股價訊息的模型。兩種資訊的結合中，與 HAN、LSTM 模型相比，最佳準確度為 80%，整體平均最高可提升 40%的準確度。相較於 Ziniu Hu[6]等人提出只有新聞資料的 HAN 模型相比，我們提出的 Fusion Model，整體平均高於 37%的準確率。本研究所提出的方法還有能改善的方向，針對股票趨勢預測中，除了本論文結合新聞文本與股價，還可以使用技術指標、情緒，增加深度學習神經網路在股票趨勢預測上的準確度。在處理新聞文章上，不是輸入整篇新聞文章，而是從新聞文章中提取對股票有影響的事件，並且把這些事件輸入到模型中進行訓練，降低多餘新聞文章中不重要的資訊，增加模型預測之精準度。

參考文獻

- [1] Indu Kumar, Kiran Dogra, Chetna Utreja, and Premlata Yadav. A comparative study of supervised machine learning algorithms for stock market trend prediction. In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pages 1003–1007. IEEE, 2018.
- [2] Kai Chen, Yi Zhou, and Fangyan Dai. A lstm-based method for stock returns prediction: A case study of china stock market. In 2015 IEEE international conference on big data (big data), pages 2823–2824. IEEE, 2015.
- [3] Heeyoung Lee, Mihai Surdeanu, Bill MacCartney, and Dan Jurafsky. On the importance of text analysis for stock price prediction. In LREC, volume 2014, pages 1170–1175, 2014.
- [4] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Using structured events to predict stock price movement: An empirical investigation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1415–1425, 2014.
- [5] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Deep learning for event-driven stock prediction. In Twenty-fourth international joint conference on artificial intelligence, 2015.
- [6] Ziniu Hu, Weiqing Liu, Jiang Bian, Xuanzhe Liu, and Tie-Yan Liu. Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In Proceedings of the eleventh ACM international conference on web search and data mining, pages 261–269, 2018.
- [7] Yuzheng Zhai, Arthur Hsu, and Saman K Halgamuge. Combining news and technical indicators in daily stock price trends prediction. In International symposium on neural networks, pages 1087–1096. Springer, 2007.
- [8] Xiaodong Li, Pangjing Wu, and Wenpeng Wang. Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong. Information Processing & Management, page 102212, 2020.
- [9] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In International conference on machine learning, pages 1188–1196, 2014.
- [10] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pages 1480–1489, 2016.
- [11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization.

arXiv preprint arXiv:1412.6980, 2014.

- [12] 參考網址來源: <https://towardsdatascience.com/stock-prediction-using-recurrent-neural-networks-c03637437578>